The International Journal of Digital Curation
Volume 8, Issue 1 | 2013

# Data Intelligence Training for Library Staff

Madeleine de Smaele, Ellen Verbakel, and Nicole Potters,
3TU.Datacentrum, Delft, The Netherlands

Marina Noordegraaf,
www.verbeeldingskr8.nl

## Abstract

The Data Intelligence 4 Librarians course was developed by 3TU.Datacentrum at the end of 2011 to provide online resources and training for digital preservation practitioners, specifically for library staff. The course objectives are to transfer and exchange knowledge about data management, and to provide participants with the skills required to advise researchers or research groups on efficient and effective ways of adding value to their data. The paper describes the process of creating the course, the methodology and the results of the first pilot, which took place from February to June 2012. It also demonstrates the choices made during the design process and discusses the implications of the evaluation of the pilot course for further development: in particular, how the course might be expanded to more disciplines and other data repositories.

*International Journal of Digital Curation* (2013), *8*(1), 218–228. http://dx.doi.org/10.2218/ijdc.v8i1.255

# Summary

The Data Intelligence 4 Librarians course was developed by 3TU.Datacentrum at the end of 2011 to provide online resources and training for digital preservation practitioners, specifically for library staff. The course is intended to overcome the insecurity and perceived lack of knowledge about data management that prevents library staff from proactively providing support to research staff with the management, storage and sharing of their research data. The course objectives are to transfer and exchange knowledge about data management, and to provide participants with the skills required to advise researchers or research groups on efficient and effective ways of adding value to their data. The course is an additional service provided by 3TU.Datacentrum, a digital repository for research data set up by the research libraries of the three Dutch Universities of Technology (3TU): Delft University of Technology, Eindhoven University of Technology, and the University of Twente. The paper describes the process of creating the course, the methodology and the results of the first pilot, which took place from February to June 2012.

There were three phases to creating the course: an investigation of the training needs, the design of the course and the development of the training materials. The training needs, as identified by library staff involved in data management, could be divided into hard skills (such as data management basics, data citation and knowledge of the data browser interface) and soft skills (such as acquisition skills, advisory skills and a reflective attitude towards their profession). This has led to a course with a competency-based modular design, consisting of four modules.

3TU.Datacentrum staff provided knowledge and experience as input for the initial design. Later on, they also provided input for one or more of the four modules. The contents of the course were written in two to four iterating sessions. The coaches were deliberately selected for their didactical skills and not as knowledge experts. The framework for the course was finished by the end of 2011.

The project team designed a blended learning course, composed of a didactical mix of group meetings, online study and homework assignments. The course website provides online reference material. The course is characterized by a practical, hands-on approach. The total study load for participants is about seven full days: four days of tuition and three days for homework assignments. By using a modern collaboration environment and social media, the course meets the needs of the modern scientific community. The training methodology proved effective in creating a solid base for digital preservation and a network of new data librarians.

In the first half of 2012, 14 librarians participated in the pilot course and provided feedback. The participants valued the course and its networking component, but felt that some of the tuition days could be more demanding. For example, actually having to write a data management plan would improve the usefulness of the training. Participants also indicated that they would appreciate hearing directly from researchers about their data management issues. The feedback has been used to improve the next cycle of the course, which started in September 2012.

This paper demonstrates the choices made during the design process and discusses the implications of the evaluation of the pilot course for further development. Finally, future plans are discussed. They include expanding the course to more disciplines and other data repositories, in cooperation with organisations from The Netherlands and the UK.

 doi:10.2218/ijdc.v8i1.255

# Introduction

Attention for research data management (RDM) from funders, high level management of universities and research institutes, and some data producers is increasing. Their goals range from improving the efficiency of research funding to ensuring research data integrity and solving practical issues around secure storage and sharing of (large) research data collections. However, the level of awareness, the available skills and the tools available for RDM differ significantly between research groups and disciplines. To bridge the gaps and support eager researchers with RDM, well-trained support staff are urgently required. In the short term, they are needed to assist data producers who are confronted with new demands, and to help young researchers reach long-term goals where RDM is integrated in the complete scientific lifecycle of (almost) all disciplines.

In the literature, emphasis lies on training researchers themselves in RDM (Lyon, 2007; Grim, Van der Heijden, De Smaele & Verbakel, 2011). During the development of our course, EDINA and Data Library, University of Edinburgh delivered MANTRA[1], an RDM training course for PhD students. To our knowledge, Data Intelligence 4 Librarians is the first course to specifically focus on support staff – particularly librarians. Library staff involved in collection development and management for a specific discipline are well-suited to fulfil RDM support tasks because of their knowledge of discipline standards, research, metadata and discovery. Data Intelligence 4 Librarians adds knowledge about digital preservation (information technology and archiving) as well as soft skills like consultancy. In our view, the continuous development of the knowledge and skills of support staff is necessary to keep up-to-date with requirements in a fast moving field. The mission of the course was formulated as follows:

> Data Intelligence 4 Librarians wants to contribute to the professionalization and positioning of support staff as a trusted partner in the support of data-intensive science.

# Process

There were three phases to designing and delivering the course *Data Intelligence 4 Librarians*: an investigation of the training needs, the design of the course, and the development of the training materials.

## Investigation of the Training Needs

The Data Intelligence 4 Librarians course was primarily intended for library staff from the three Universities of Technology in The Netherlands (3TU): Delft University of Technology, Eindhoven University of Technology, and the University of Twente. The first step was a visit to Rob Grim, the only person in The Netherlands at that time with the official job title "data librarian", to interview him on training needs for data librarians-to-be. Together with colleagues from 3TU.Datacentrum[2] and

[1] MANTRA – Research Data Management Training: http://datalib.edina.ac.uk/mantra/
[2] 3TU.Datacentrum: http://datacentrum.3tu.nl

NIOO/KNAW[3] he was writing a white paper stressing the urgency for the development of RDM training for support staff and researchers. The white paper described the results of a literature search on curation education, which showed that there were no examples that could be transferred to the current situation in The Netherlands (Grim, Van der Heijden, De Smaele & Verbakel, 2011).

Given such a starting point, information sessions were organised with our target group: the information specialists or reference librarians of the 3TU universities. In the course of a full day, current knowledge on data management and 3TU.Datacentrum was shared, and training needs were identified during break-out sessions. Building on the feedback, a second session was organised at Delft University of Technology for a wider audience, including acquisition managers and project/support staff, in addition to the information specialists. The results of all sessions were compared and the similarities were significant. Library staff wanted to learn more about technical details of data management, such as data formats, the software researchers use, metadata, data processing and file sharing systems, but also about database law, data and copyright, and research methods. Additionally, they wanted to develop soft skills like networking, interview techniques, convincing and inspiring researchers, and mobilising and coordinating efforts. Interviewees were uncertain about how far their subject-specific knowledge should reach in order to be of value to their customers. Uncertainty about the questions they might be asked and the answers they should give in conversations with researchers hinders good contact with those researchers. Nevertheless, the interviewees stated that good contact with some researchers was already established.

The interviewees were aware that they should acquire good knowledge of current developments in the data management field in The Netherlands, Europe and worldwide. Current knowledge would help to identify which paths should be followed. Interviewees indicated that they wanted to learn more about using Web 2.0 applications as a means of keeping up with current developments. In another two-hour meeting, the identified training needs were shared with two of the information specialists at Eindhoven University of Technology, who had experience in data curation services. They agreed upon the goals formulated by their colleagues.

To reach outside the domain of the three technical universities, the Twitter account @datalibrarians[4] was set up to proactively look out for the latest developments in the field. To get in touch with relevant parties in The Netherlands, the Onderzoeksdataforum (Research Data Forum) – now the Special Interest Group Research Data[5] – was visited, and in June 2011 a visit was paid to the ICE-forum[6] to get up-to-date with the latest developments in data curation education. Our pragmatic approach received confirmation: get the course running as soon as possible and let participants evaluate its content to improve it. Open Educational Resources on the subject were not yet available at that time. To prevent other parties from having to reinvent the wheel, plans were made to put the contents of our course online.

[3] NIOO/KNAW: http://www.nioo.knaw.nl/en/

[4] 3TU.Datacentrum on Twitter: http://www.twitter.com/datalibrarians

[5] Special Interest Group Research Data: https://www.surfspace.nl/sig/28-research-data/29-over-de-sig/

[6] International Curation Education Forum: http://www.jisc.ac.uk/whatwedo/programmes/preservation/iceforum.aspx

 doi:10.2218/ijdc.v8i1.255

## Design of the Course

During the investigation of training needs, librarians expressed a great need for more knowledge and ICT-skills before they would feel capable of establishing data services. Although specialist knowledge and skills are required when providing data services, the gap is not simply a knowledge and skills gap. It is as much an attitude gap. Are people willing to proactively seek out the knowledge they need? Are they confident enough to not know everything when engaging in a conversation with a researcher? Do they possess the necessary skills to both convey their message to a researcher and ask the necessary questions to add to their knowledge base? In competency-based education, students can integrate knowledge, skills and attitude and put the three into practice.

Our course is designed and built as a competency-based modular course, combining online and face-to-face tuition (blended learning). From the inventory of learning goals, seven core competencies were defined, as shown in Table 1. The seven competencies were intended as directional in the development of the course, rather than goals in themselves.

| Core competency | Learning Goals |
|---|---|
| Skilfully handles ICT | • Uses the available Information Technology in an effective and efficient way.<br>• Can use the 3TU.Databrowser to upload a dataset and make it available for (re)use. |
| Has specific library knowledge | • Knows how to acquire specific knowledge about metadata standards.<br>• Can explain how minting a DOI (Digital Object Identifier) and UUID (Universally Unique Identifier) enhances the visibility (citability) of a dataset. |
| Develops entrepreneurship | • Is committed to improving data services in response to changing needs in the field.<br>• Keeps an eye on trends in the profession, knows where new knowledge may be found (networks) and spreads relevant information to key persons in the organisation.<br>• Investigates needs in the field by means of questionnaires, interviews and so-called focus groups.<br>• Actively contributes to developments in the field by visiting conferences and enrolling in courses and training. |
| Develops a systemic view | • Acknowledges that data are just one part of the scientific research cycle and is aware of the significance of data within that cycle.<br>• Sees the library and its data and information services as part of a larger decision-making system. |

| Core competency | Learning Goals |
|---|---|
| Develops advisory skills | • Can advise researchers on RDM topics, such as sustainable data formats, data models, intellectual property and the demands of research funders.<br>• Knows when certain aspects of RDM do not fit his/her expertise and is able to refer questions to corresponding knowledge experts.<br>• Supports researchers in setting up a data management plan (DMP).<br>• Can hold a so-called data interview and is aware of the possible use of data curation profiles (DCP) as a possible interview instrument.<br>• Can connect to a researcher's perception on data management and data sharing.<br>• Asks for feedback on his/her advisory skills and adjusts his/her behaviour accordingly. |
| Develops collaboration skills | • Investigates how collaborating with other employees, institutions, data centers and researchers may lead to a better provision of services.<br>• Acknowledges the need for a forum of data professionals who may join forces in important data issues, such as copyright and (inter)national information infrastructure.<br>• Takes responsibility for his/her role in partnerships. |
| Develops training materials | • Develops RDM training materials for end users.<br>• Is able to translate the knowledge and skills acquired in the Data Intelligence 4 Librarians training into RDM training material for different target audiences. |

Table 1. The seven core competencies of the Data Intelligence 4 Librarians course.

Using these competencies, a modular course was designed which consists of the following modules:

1. Current Topics
2. Data Management
3. Technical Skills
4. Advisory Skills

Table 2 maps the four modules with the competencies worked on during each particular module.

 doi:10.2218/ijdc.v8i1.255

| Module | Competencies[7] |
|---|---|
| Current Topics | • Develops entrepreneurship<br>• Develops a systemic view<br>• Develops collaboration skills |
| Data Management | • Develops advisory skills<br>• Has specific library knowledge |
| Technical Skills | • Skilfully handles ICT<br>• Has specific library knowledge |
| Advisory Skills | • Develops advisory skills<br>• Develops a systemic view<br>• Develops entrepreneurship<br>• Develops collaboration skills |

Table 2. Identified Modules and Competencies.

For the Current Topics module, the homework assignment requires the students to use relevant (online) tools and sources for about a month to form a opinion on the current state of affairs in the field of data curation (services). In a short elevator pitch, they then share their findings with their fellow students.

The premise of the course is that supporting researchers in data curation involves teamwork. A team of professionals who have insight into each other's knowledge and experience, thereby enabling them to give researchers the best support. As a consequence, strong emphasis is placed on enabling networking within the student group. Meeting fellow students who work in the same field and encounter similar or different RDM dilemmas stimulates people to find solutions. The online component of the course consists of the course website,[8] as well as the use of social media (such as Twitter and Google+) which encourages students to put the networking component of the course into practice.

The Current Topics module enables students to feel that they are in control of collecting the information they need, and that they can find and connect to the people who have the answers they need. Without developing entrepreneurship, knowledge becomes obsolete very fast, which makes it difficult to provide effective data services.

In the second module, students learn about research data management in general. Where do research data fit into the research lifecycle? Which data formats exist? Which data are suitable for depositing into a data archive and why? Which metadata

[7] The competency 'Develops training materials' is not listed in Table 2. It is meant to be developed after completing the Data Intelligence 4 Librarians course. Once support staff have acquired knowledge and skills in data curation, they have a solid base to subsequently develop training materials for the students and researchers they support.

[8] Data intelligence: http://dataintelligence.3tu.nl

should be added to ensure that a future user understands what the dataset comprises? What is the current policy regarding data management within research institutions? How can you assist a researcher in writing a data management plan? What about copyright?

The third module is technical in nature. It explains what digital objects are, how data are to be cited, how research data are transformed to different formats during different phases in the lifecycle, how to search for data (depending on the data model used) and how to enhance publications. Students practice searching for existing datasets, and describing and uploading a dataset.

For the fourth module, students write an acquisition plan. They present their plan to fellow students and give feedback on each others' plans. They try their advisory skills in role plays and, finally, put their acquisition plan into practice by actually carrying out the first steps described in their own plans. In this final module theory, skills and attitude come together.

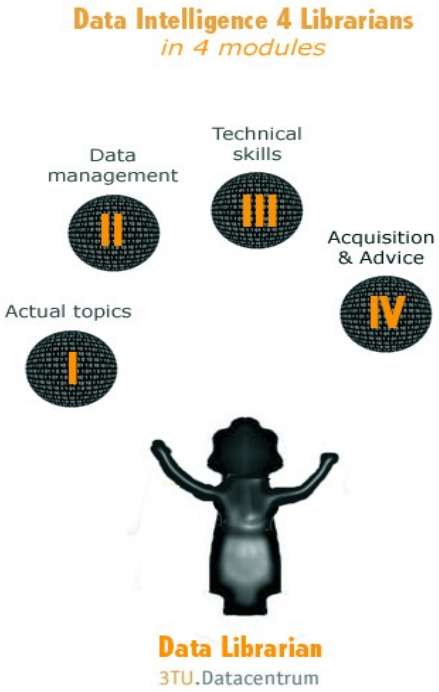


Figure 1. A graphic overview of the course: A data librarian juggling four balls.[9]

## Development of Training Materials

Once the training design was finished, the actual development of the course started. For each module a team of three to four experts on the topic came together for two to four iterating sessions of one or two hours. The content was written in between sessions. The design was fine-tuned with the feedback of knowledge experts, who reviewed and re-wrote the texts. The experts agreed on the final text used on the course website and on the homework assignments. As part of the Acquisition & Advice module, four cases detailing actual experiences of library staff in the field of

[9] Data Intelligence for Librarians: http://dataintelligence.3tu.nl/

 doi:10.2218/ijdc.v8i1.255

data management were included. The whole process of developing the training material for all four modules took no more than three months. Even though the course was initially designed for Dutch participants, the decisions was made to translate the course website into English, in order to provide a possible source of inspiration for universities or other institutions abroad. The course material was illustrated with specially created images. One contact person from each of the 3TU universities performed a final check on the web content.

Two coaches were selected for their didactical skills, following the belief that being an inspiring coach requires different skills than being a knowledge expert. The design of the course places much emphasis on learning by trying things yourself, without putting too much trust in an expert. Everyone is an expert in a certain area and data curation is teamwork where everyone should be valued for their specific contribution.

The coaches were provided with a global scenario for each training day, homework assignments, and some PowerPoint presentations not included in the course website.

## Findings

The pilot training was mainly attended by library staff from the 3TU universities. On the final day of the course the participants were asked to fill in a course evaluation form. The coaches also wrote an evaluation report. The feedback was used to improve the course for the next cycle, which was held from September to December 2012. Some important issues put forward by participants and coaches were:

- The setup of the course was not completely clear to everyone. An important aspect is the exchange of knowledge and experiences between the participants themselves. Some participants indicated that they had expected more traditional lectures.
- Participants enjoyed and appreciated the discussions resulting from the homework assignments. These exchanges of ideas and information were seen as the most valuable element of the course.
- Four days of face-to-face tuition were seen as a considerable time investment, but useful because of the relevant discussions and networking possibilities.
- The participants' knowledge of tools like Google+ and Google Drive was less than expected. It would be advisable to provide more support for these tools.
- Participants were interested to hear from researchers themselves how they deal with data management issues, and about differences between disciplines.
- Participants would have liked more actual use cases, preferably presented by an inspiring, enthusiastic person.
- Participants missed the opportunity to practice writing an actual Data Management Plan.
- Participants urgently needed practical information about setting up a front office for data management services.
- The participants appreciated the images included in the course material on the website, which they thought were a memorable way to clarify concepts.

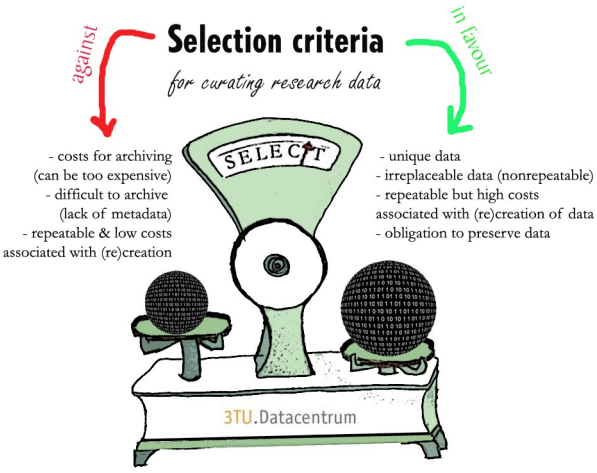


Figure 2. One of the illustrations used on the course website.[10]

## Future

In view of the evaluation results, a collaboration with DANS (Data Archiving and Networked Services – an archive of humanities, archaeology, geospatial sciences and behavioural and social sciences in The Netherlands) was initiated to expand the course to other disciplines. The original setup of the course and methodology remained intact, while the learning materials were expanded to include information about non-technical disciplines as well as services provided by DANS. DANS also provided one of the coaches for the second cycle of the course, making it a truly joint undertaking.

The second cycle of the course took place between September and December 2012 with 16 participants from universities and research organisations throughout The Netherlands. First impressions are that the diversity of the participants, who come from a wider range of organisations and disciplines than in the first cycle, adds value to the group discussions and knowledge exchange during the course. The social media component of the course also seems to work better, possibly because most participants did not know each other before the course and were more reliant on social media to exchange information. Of course, this second cycle will also be evaluated and the evaluations will be used for further improvements.

As information about the Data Intelligence 4 Librarians course has spread within and even outside The Netherlands, several parties have asked to share our experience and knowledge in other projects, including:

- RDMRose,[11] "a JISC-funded project to produce taught and continuing professional development (CPD) learning materials in Research Data Management (RDM) tailored for information professionals."
- Exchange experience and work with UK partners, such as the Digital Curation Centre and JISC.

[10] Retrieved from http://dataintelligence.3tu.nl/

[11] The University of Sheffield Information School: http://www.sheffield.ac.uk/is/research/projects

Our Data Intelligence 4 Librarians course has also inspired the development of the Liaison Librarian Training by EDINA and the Data Library at the University of Edinburgh (Macdonald, Donnelly, & Rice 2012).

Participants in the course have expressed appreciation for the visual content on the course website, and adding more audio-visual content will greatly enhance the learning experience. In 2013 the website will be updated and get an audio-visual makeover. Simultaneously, an RDM training for researchers will be developed, initially for use at the three Universities of Technology in The Netherlands (3TU).

## Conclusions

The development of the Data Intelligence 4 Librarians course was a leap into the unknown. No courses on this topic existed in The Netherlands and no easily reusable courses on this topic existed abroad. However, the need to get such a course running was obvious. Our pragmatic approach of just getting started has paid off. Participants, while appreciating the course as a whole, have indicated some of its shortcomings, thereby enabling improvements.

In the second cycle of the course, much of the participants' feedback has been incorporated and it is now spreading from the three technical universities to a broader audience. With Data Intelligence 4 Librarians, a course has been developed which is suitable for an even broader scope, depth and audience than initially intended.

## Acknowledgements

The authors thank Jeroen Rombouts and Alenka Princič for their contributions to this paper.

## References

Grim, R., Van der Heijden, M., De Smaele, M. & Verbakel, E. (2011). Witboek dataprofessionals in Nederland. White Paper Data Professionals in The Netherlands. Retrieved from http://www.surf.nl/nl/publicaties/pages/witboekdataprofessionals.aspx

Lyon, L. (2007). Dealing with Data: Roles, Rights, Responsibilities and Relationships. UKOLN. Retrieved from http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing_with_data_report-final.pdf

Macdonald, S., Donnelly, A. & Rice, R. (2012). *JISC Managing Research Data: Liaison Librarian Training.* Retrieved from http://www.slideshare.net/edinadocumentationofficer/jisc-managing-research-data-liaison-librarian-training